

Heats of Formation of Organic Molecules Calculated by Density Functional Theory: II. Alkanes

JAN LABANOWSKI,¹ LAWRENCE SCHMITZ,² KUO-HSIANG CHEN,³
NORMAN L. ALLINGER³

¹Ohio Supercomputer Center, 1224 Kinnear Road, Columbus, Ohio 43212-1163

²Department of Chemistry, Marshall University, Huntington, West Virginia 25755

³Computational Center for Molecular Structure and Design, Department of Chemistry, University of Georgia, Athens, Georgia 30602-2526

Received 6 June 1997; accepted 10 March 1998

ABSTRACT: Heats of formation of alkanes have been calculated with an accuracy of better than 0.36 kcal/mol by using the total energy calculated by density functional theory, plus bond and group equivalents and statistical mechanical corrections. The necessary equivalents were assigned to bonds and groups in molecules. Once such equivalents have been derived from the fit to available experimental values for a large and diverse set of compounds, they can be used to predict heats of formation for compounds of the same class for which these quantities are not experimentally available. Expanding the method to a new class of compounds requires that only new groups of equivalents for that class be added to the scheme. This provides a path for the systematic expansion of the model to new classes of compounds, and gives us a computational method for getting around the lack of experimental information about systems of interest. © 1998 John Wiley & Sons, Inc. *J Comput Chem* 19: 1421–1430, 1998

Keywords: heat of formation; density functional theory; bond equivalents; group equivalents

Introduction

Heats of formation are the fundamental quantities representing the energies of compounds in their standard states. For this reason,

they are widely useful for a variety of purposes. With the recent focus on energy efficiency, and on the environmental impact of combustion, these data have received even more popular interest. Unfortunately, however, these quantities are quite difficult to accurately measure experimentally. Moreover, such calorimetric studies are very time

consuming, and require the utmost care with respect to the design and conditions of the measurements. For these reasons, only a few experimental results along these lines have been reported in recent years. Thus, there is a real need to develop reliable computational models that would allow accurate predictions of these quantities.

Heats of formation can be calculated *de novo* by using the currently most advanced methods of quantum chemistry. However, such calculations can only be routinely carried out for small molecules containing a few atoms. An example of research leading in this direction is the compilation of heats of atomization calculated with the G2 approach,¹ where for a series of 55 small molecules (up to three nonhydrogen atoms), the average error reported was 1.6 kcal/mol. However, the fact that fourth order perturbational theory and quadratic CI approaches were used in the study suggest that such an accuracy in *de novo* calculations of heats of formation will not be available soon for routine applications to large molecules.

We take another approach here, which is to carry out less time-consuming quantum calculations, and then adding corrections in a physically reasonable way so as to fit the available database of reliable, experimentally derived, heats of formation. One might hope that such methods would enable accurate prediction of heats of formation. Of course, the test is to actually make such predictions, and check their accuracy. This approach is based on the idea of isodesmotic reactions, and has been used several times previously. However, for reasons outlined in what follows it seemed necessary to redo these calculations.

Heats of formation have long been calculated by "bond energy" schemes.² In more recent years, molecular mechanics has been used.³ The basic method is the same using either molecular mechanics or quantum mechanics, and has been reviewed.⁴

The original idea of using quantum mechanical calculations in this way was put forth independently by Schleyer⁵ and by Wiberg,⁶ and applied by them to a representative test set of structures. Although their results were not experimentally accurate, they were sufficiently good to suggest the method would be worth refining, and that one might reasonably hope to obtain experimental accuracy in this way. We subsequently developed the method using Hartree–Fock energies calculated at the 6-31G* level with rather good success (for refs. see 4). However, the lack of correlation energy in this approach suggests that certain sys-

tems, such as conjugated molecules, would probably not be well treated and, in general, the accuracy would be suspect. However, to accurately account for correlation brings us to something like the G2 approach, which is currently not feasible for molecules of the size of interest. The idea of using isodesmotic reactions with empirical parameters to deal with what could not be accurately calculated appeared to be the only option. Accordingly, we previously explored the possibility of using DFT calculations,⁷ with every indication that this method would be superior to the Hartree–Fock method, and could be extended to more diverse systems with a higher degree of confidence.

Computational Methods

The details of the calculations are described in our earlier study⁷ on heats of formation of alkanes. In our model, the approximation to the heat of formation (ΔH_f) is calculated from the following equation:

$$\Delta H_f = E_{DFT} + \sum_j a_j n_j + POP + T/R \quad (1)$$

where E_{DFT} is the DFT energy for the lowest energy conformer, a_j are the bond/group equivalents needed to correct the E_{DFT} , n_j represent the counts of groups/bonds, and POP and T/R denote statistical mechanical corrections:

- *POP*: Boltzmann correction accounting for low-energy conformations populated at room temperature. These were estimated using molecular mechanics.
- *T/R*: This term ($4RT$) accounts for translational and rotational degrees of freedom for nonlinear molecules ($6 \times 1/2 RT$) and an additional RT to convert from energy to enthalpy.
- *TOR* The number of bonds in the molecule with low-energy rotational barriers (< 5 kcal/mol), excluding methyl groups.

Note that *TOR* is taken as an adjustable parameter in this study (one of the n_j). It is to allow for the extra enthalpy in a molecule as a result of low torsional barriers, and is expected from the work of Pitzer and Gwinn⁸ to have a numerical value of about 0.4 kcal/mol. The justification and derivation details for these corrections were given in our earlier study.⁷ When the needed equivalents for

groups and bonds are known for a given molecule, the heat of formation can be easily calculated by adding these additional terms to the DFT energy of the molecule.

The equivalents, a_j , correcting the value of DFT electronic energy, E_{DFT} , are derived from the through-origin multiple regression carried out on the set of experimental heats of formation $\Delta H_f[i]$:

$$Y[i] = \sum_j a_j n_j[i] \quad (2)$$

where $Y[i] = \Delta H_f[i] - E_{\text{DFT}}[i] - \text{POP}[i] - T/R - \sum_k a_k^* n_k[i]$. In this expression, $n_j[i]$ is the count of the j th group or bond in the i th molecule. The a_k^* are the previously defined equivalents, in the present case, for alkanes (see Table I).⁷ This scheme works well in general, provided that the groups corresponding to parameters a_j were carefully chosen, and expanding the model to compounds with a new functional group/substituent does not require refitting equivalents already derived. Obviously, the same basis sets, grid density, and fitting functions have to be used in all of the quantum calculations for the equivalents to be transferable.

We previously reported a DFT study as outlined for alkanes,⁷ and initially started with the increments determined in that study. When we studied larger molecules than had previously been examined (up to 10 carbon atoms) we noticed systematic errors creeping in, which became increasingly worse with molecular size. It was apparent that the data set used earlier to determine the group equivalents, being limited to relatively small molecules (C6 and less), was part of the problem. Large molecules contain more bonds, and more equivalents, and therefore more errors when these things are summed. Our original small molecule set contained only a few tertiary and quaternary carbon arrangements. This statistical distribution of different groups produced somewhat biased values of equivalents, which surfaced when we applied these values to larger alkanes. It became clear that we needed to use a more representative set of basis compounds, and to include larger molecules in this set. We also re-examined the Hartree–Fock calculations, because if they are to be useful anywhere, it would be with alkanes. If the correlation in alkanes can be completely taken up in the group equivalents, then the results by the Hartree–Fock method should be essentially the same as by the DFT method. However, using earlier calculations as a basis, we found that the

Hartree–Fock calculations were somewhat poorer. Even after the problem of dealing with larger molecules was taken into account, the results were still not very satisfactory. An additional problem was that we had taken much of the Hartree–Fock data from the literature, rather than doing the calculations ourselves, because these early beginnings were carried out years ago when computational power was much less available. What we found was that the earlier calculations reported in the literature were based on structures that were not always optimized adequately. There were arbitrary errors in the reported Hartree–Fock energies of up to 1 kcal/mol in many of the reported data, and these errors introduced additional unwanted scatter in the results.

To get around these two problems of inaccurate literature data and inadequate distribution of molecules in the test set of compounds, we chose a new and larger set, and carried out the computations on all of these compounds by both HF and DFT methods. The set of compounds chosen is listed in Table I. This set is adapted from similar sets used earlier for molecular mechanics heat of formation calculations,^{9,10} and the individual compounds were chosen because they are exceedingly diverse, representing all kinds of saturated hydrocarbons (excluding only three- and four-membered rings), and they are compounds for which experimental data are available, and believed to be reliable. In each case, our criterion for the convergence of the HF energy value was that the maximum residual force was reduced below 0.00045 Hartree/Bohr.

Seven parameters need to be determined to define these compounds. They are the number of CC bonds, the number of CH bonds, the number of rotatable bonds (TOR), and the number of each of the structural units, methyl, iso, and neo. Two additional parameters are used, R5 and R6. It was previously found in molecular mechanics work⁹ that an additional parameter is required, which is the number of rings present in the molecule. It was also found that, with both MM3⁹ and MM4,¹⁰ the heats of formation can be more accurately calculated if one includes a special parameter for the five-membered rings present in the molecule (and additionally, one needs special parameters for three- and four-membered rings, if those kinds of structures are included). Physically, this means that the bonding in three-, four-, and five-membered rings is different from the bonding in larger rings (six-membered rings and larger are all equivalent in this respect), and different from each other,

TABLE I.
Input Data for Alkanes (RHF / 6-31G* and 3LYP / 6-31G*).

Compound	POP ^a	Total energy (HF / 6-31G*)	Total energy (B3LYP / 6-31G*)	C—C	C—H	ME	Iso	Neo	TOR	R6	R5
Methane	0.00	−40.19517	−40.51839	0	4	0	0	0	0	0	0
Ethane	0.00	−79.22876	−79.83045	1	6	2	0	0	−1	0	0
Propane	0.00	−118.26365	−119.14425	2	8	2	0	0	0	0	0
<i>n</i> -Butane	0.30	−157.29841	−158.45800	3	10	2	0	0	1	0	0
<i>n</i> -Pentane	0.65	−196.33310	−197.77170	4	12	2	0	0	2	0	0
<i>n</i> -Hexane	1.00	−235.36779	−237.08546	5	14	2	0	0	3	0	0
<i>n</i> -Heptane	1.35	−274.40249	−276.39923	6	16	2	0	0	4	0	0
<i>n</i> -Octane	1.70	−313.43718	−315.71295	7	18	2	0	0	5	0	0
<i>n</i> -Nonane	2.05	−352.47187	−355.02667	8	20	2	0	0	6	0	0
<i>n</i> -Decane	2.40	−391.50657	−394.34038	9	22	2	0	0	7	0	0
Isobutane	0.00	−157.29898	−158.45888	3	10	3	1	0	0	0	0
Isopentane	0.07	−196.33181	−197.77116	4	12	3	1	0	1	0	0
Neopentane	0.00	−196.33383	−197.77299	4	12	4	0	1	0	0	0
Cyclopentane	0.00	−195.16358	−196.55709	5	10	0	0	0	1	0	1
Cyclohexane	0.00	−234.20801	−235.88063	6	12	0	0	0	0	1	0
Cycloheptane	0.00	−273.23036	−275.18346	7	14	0	0	0	1	1	0
Cyclooctane	0.00	−312.25650	−314.49038	8	16	0	0	0	0	1	0
Cyclononane	0.57	−351.28388	−353.79818	9	18	0	0	0	0	1	0
Cyclodecane	0.37	−390.31683	−393.11139	10	20	0	0	0	0	1	0
Cyclododecane	0.41	−468.39383	−471.74639	12	24	0	0	0	0	1	0
2,4-Dimethylpentane	0.18	−274.39961	−276.39773	6	16	4	2	0	2	0	0
3-Ethylpentane	0.00	−274.39588	−276.39457	6	16	3	1	0	3	0	0
2,2,3-Trimethylbutane	0.00	−274.39361	−276.39319	6	16	5	1	1	1	0	0
2,2-Dimethylbutane	0.00	−235.36461	−237.08370	5	14	4	0	1	1	0	0
2,3-Dimethylbutane	0.06	−235.36306	−237.08252	5	14	4	2	0	1	0	0
3,3-Dimethylpentane	0.00	−274.39470	−276.39391	6	16	4	0	1	1	0	0
2,2-Dimethylpentane	0.00	−274.39897	−276.39712	6	16	4	0	1	2	0	0
Di- <i>t</i> -butylmethane	0.02	−352.45511	−355.01420	8	20	6	0	2	2	0	0
3,3-Diethylpentane	0.00	−352.45589	−355.01499	8	20	4	0	1	0	0	0
2-Methylhexane	0.41	−274.40105	−276.39846	6	16	3	1	0	3	0	0
3-Methylhexane	0.39	−274.39881	−276.39680	6	16	3	1	0	3	0	0
2,2-Dimethylhexane	0.06	−313.43366	−315.71088	7	18	4	0	1	3	0	0
2,3-Dimethylhexane	0.31	−313.43021	−315.70815	7	18	4	2	0	3	0	0
2-Methylheptane	0.66	−313.43574	−315.71217	7	18	3	1	0	4	0	0
3-Methylheptane	0.54	−313.43351	−315.71053	7	18	3	1	0	4	0	0
4-Methylheptane	0.49	−313.43332	−315.71043	7	18	3	1	0	4	0	0
2,2,3,3-Tetramethylbutane	0.00	−313.42007	−315.70101	7	18	6	0	2	0	0	0
2,2,3,3-Tetramethylpentane	0.00	−352.44971	−355.01038	8	20	6	0	2	0	0	0
3-Ethyl-3-methylpentane	0.13	−313.42566	−315.70467	7	18	4	0	1	3	0	0
Methylcyclopentane	0.10	−234.19967	−235.87257	6	12	1	1	0	1	0	1
Ethylcyclopentane	0.30	−273.23339	−275.18547	7	14	1	1	0	2	0	1
1,1-Dimethylcyclopentane	0.27	−273.23297	−275.18552	7	14	2	0	1	1	0	1
Norbornane	0.00	−272.06120	−273.96849	8	12	0	2	0	0	0	2
<i>cis</i> -Bicyclo[3.3.0]octane	0.00	−311.10396	−313.29048	9	14	0	2	0	1	0	2
<i>trans</i> -Bicyclo[3.3.0]octane	0.00	−311.08985	−313.27764	9	14	0	2	0	0	0	2
Protoadamantane	0.00	−388.00723	−390.70785	12	16	0	4	0	0	2	1
eq-Methylcyclohexane	0.10	−273.24366	−275.19563	7	14	1	1	0	0	1	0
Bicyclo[2.2.2]octane	0.00	−311.10360	−313.28994	9	14	0	2	0	0	2	0
Adamantane	0.00	−388.02648	−390.72520	12	16	0	4	0	0	3	0
<i>trans</i> -Decalin	0.00	−389.18906	−391.93368	11	18	0	2	0	0	2	0
<i>cis</i> -Decalin	0.00	−389.18357	−391.92852	11	18	0	2	0	0	2	0
Tri- <i>t</i> -butylmethane	0.00	−508.52125	−512.21042	12	28	9	1	3	0	0	0

^a Units for POP and total energy are kcal / mol and a.u., respectively.

so that the zero of energy for one of these small rings has to be established with the aid of a parameter. We had previously thought that no such parameter would be necessary in quantum mechanical calculations of the heats of formation, as this would be taken care of in the calculation of the total energy of the system (E_{DFT}) automatically; however, that proves not to be the case.

The number of bonds that have low rotational barriers is needed (*TOR*), together with the energy increase due to the Boltzmann distribution of higher energy conformations (*POP*). The values *POP* and E_{DFT} are given in Table I.

Tables II and III give the derived group and bond equivalents, and the calculated heats of formation for a number of alkanes determined for the HF method. Analogous data for the DFT method are given in Tables IV and V.

The first step in these calculations always involved the geometry optimization of the molecule under study by the HF method, with Gaussian-94.¹¹ Starting molecular geometries were usually taken from our previous molecular mechanical calculations.

It should be possible to carry out still better quantum mechanical calculations by either increasing the size of the basis set, or by including electron correlation. Improving basis set size would be computer-time intensive, and probably would not help very much. The empirical heat parameters would tend to adsorb the systematic errors, and these would be what one would expect to be improved upon most by increasing the basis set size.

We used another approach. The Hartree-Fock-derived geometries were reoptimized with the DFT calculations. Usually only a few iterations were needed to reach the minimum. The resulting DFT energies should include an approximation to the correlation energy. How well this approach worked is presented next.

TABLE II.
Group and Bond Equivalents for Alkanes from
RHF/6-31G* Calculation.

C—C	18.943089 a.u.	(11887.167 kcal/mol)
C—H	10.040710 a.u.	(6300.747 kcal/mol)
ME	0.004173 a.u.	(2.619 kcal/mol)
Iso	−0.005558 a.u.	(−3.488 kcal/mol)
Neo	−0.013576 a.u.	(−8.519 kcal/mol)
<i>TOR</i>	0.001594 a.u.	(1.000 kcal/mol)
R6	0.009528 a.u.	(5.979 kcal/mol)
R5	0.007327 a.u.	(4.598 kcal/mol)

We used two different implementations of the DFT method: Gaussian-94 with the B3LYP¹¹ and a 6-31G* basis set, and deMon with the classic Becke/Perdew gradient corrections,¹² as in our previous study.⁷ They differ not only in GGA corrections used,[‡] but also in the way the integration is done. Obviously, because different gradient corrections for exchange and correlation are used, the energies from Gaussian and deMon are different. The advantage of the deMon program was that it was much faster than corresponding Gaussian-94 calculations. However, we did not use the latest version of deMon (our version was from 1994), and for larger molecules we occasionally noticed strong dependence of DFT energy on the starting orientation of the molecule. For identical internal geometries, the DFT energies for different orientations differed for some larger molecules in excess of 1 kcal/mol. This computational artifact is known to be related to the grid density and the number of spherical grid shells around atoms. However, this version of deMon did not allow denser grids and additional grid shells. For problem cases we ran the calculations for a few different orientations of the given molecule and chose the one that resulted in the lowest energy.

Most results from deMon calculations were satisfactory. However, for reasons described in what follows, we decided not to base conclusions on the deMon results, but rather turned back to Gaussian calculations. The first stage of the Gaussian calculations was geometry optimization at the Hartree-Fock level. The resulting HF/6-31G* energies were used to derive a set of group and bond equivalents for the HF level of calculations. The HF calculations were followed by DFT (B3LYP) geometry optimization. This step required only a few iterations for most molecules, and therefore did not add substantially to the computational expense. The resulting DFT energies were then used to derive a new set of group and bond equivalents for the DFT level of calculations.

As mentioned earlier, we believe that the deMon results would be as good as those obtained from Gaussian, provided that better grids are used. Such options may be available with more recent, or future versions of deMon, and then it would not matter which program were used for deriving equivalents. We did not study the matter further, however.

Gaussian is not free from occasional glitches, either. In cases where we encountered an outlier,

[‡]GGA (generalized gradient approximation).

TABLE III.
Calculated Heats of Formation for Alkanes by RHF / 6-31G* Calculation.^a

Compound	WT ^b	H _f ⁰ _{calc.}	H _f ⁰ _{exp.}	Δ _{calc.-exp.}
Methane	1	-17.89	-17.89	0.00
Ethane	6	-19.35	-20.04	0.69
Propane	9	-24.87	-25.02	0.15
<i>n</i> -Butane	8	-29.99	30.03	0.04
<i>n</i> -Pentane	7	-35.03	35.08	0.04
<i>n</i> -Hexane	7	-40.08	-39.95	-0.13
<i>n</i> -Heptane	6	-45.12	-44.89	-0.23
<i>n</i> -Octane	5	-50.16	-49.82	-0.34
<i>n</i> -Nonane	4	-55.20	-54.75	-0.45
<i>n</i> -Decane	2	-60.25	-59.64	-0.61
Isobutane	9	-32.52	-32.07	-0.45
Isopentane	7	-36.68	-36.92	0.24
Neopentane	9	-41.43	-40.14	-1.29
Cyclopentane	9	-17.75	-18.44	0.69
Cyclohexane	8	-29.88	-29.50	-0.38
Cycloheptane	7	-27.52	-28.22	0.69
Cyclooctane	7	-29.54	-29.73	0.19
Cyclononane	6	-30.77	-31.73	0.95
Cyclodecane	4	-36.27	-36.88	0.61
Cyclododecane	0	-53.80	-54.58	0.78
2,4-Dimethylpentane	6	-48.22	-48.21	-0.01
3-Ethylpentane	6	-44.20	-45.25	1.06
2,2,3-Trimethylbutane	6	-48.05	-48.95	0.90
2,2-Dimethylbutane	5	-44.36	-44.48	0.12
2,3-Dimethylbutane	7	-41.79	-42.49	0.70
3,3-Dimethylpentane	6	-47.87	-48.08	0.21
2,2-Dimethylpentane	6	-49.54	-49.20	-0.34
Di- <i>t</i> -butylmethane	5	-57.28	-57.80	0.52
3,3-Diethylpentane	7	-56.50	-55.67	-0.83
2-Methylhexane	5	-47.03	-46.52	-0.51
3-Methylhexane	5	-45.65	-45.73	0.08
2,2-Dimethylhexane	5	-54.91	-53.68	-1.23
2,3-Dimethylhexane	5	-50.92	-51.10	0.18
2-Methylheptane	5	-52.16	-51.47	-0.70
3-Methylheptane	5	-50.88	-50.79	-0.10
4-Methylheptane	5	-50.82	-50.66	-0.16
2,2,3,3-Tetramethylbutane	5	-52.69	-53.92	1.23
2,2,3,3-Tetramethylpentane	5	-55.91	-56.64	0.73
3-Ethyl-3-methylpentane	5	-49.78	-51.35	1.57
Methylcyclopentane	2	-25.79	-25.27	-0.52
Ethylcyclopentane	5	-30.37	-30.34	-0.03
1,1-Dimethylcyclopentane	6	-33.55	-33.04	-0.51
Norbornane	5	-12.93	-12.42	-0.52
<i>cis</i> -Bicyclo[3.3.0]octane	3	-23.39	-22.07	-1.32
<i>trans</i> -Bicyclo[3.3.0]octane	4	-15.53	-15.86	0.33
Protoadamantane	2	-19.35	-20.53	1.19
eq-Methylcyclohexane	9	-37.64	-36.99	-0.64
Bicyclo[2.2.2]octane	4	-21.40	-22.58	1.18
Adamantane	2	-30.05	-31.76	1.71
<i>trans</i> -Decalin	6	-44.28	-43.54	-0.72
<i>cis</i> -Decalin	6	-40.82	-40.45	-0.37
Tri- <i>t</i> -Butylmethane	0	-43.44	-56.40	2.98

Standard deviation = 0.71 kcal / mol; weighted standard deviation = 0.67 kcal / mol.

^a In kcal / mol.^b Weights were determined according to the accuracy of experimental data and overall importance of the compound.

TABLE IV.
Group and Bond Equivalents for Alkanes from
B3LYP / 6-31G* Calculation.

C—C	19.061549 a.u.	(11961.504 kcal / mol)
C—H	10.121515 a.u.	(6351.453 kcal / mol)
ME	0.002633 a.u.	(1.652 kcal / mol)
Iso	−0.004287 a.u.	(−2.690 kcal / mol)
Neo	−0.010863 a.u.	(−6.816 kcal / mol)
TOR	0.000727 a.u.	(0.456 kcal / mol)
R6	0.002186 a.u.	(1.372 kcal / mol)
R5	0.001118 a.u.	(0.702 kcal / mol)

TABLE V.
Calculated Heats of Formation for Alkanes by
B3LYP / 6-31* Calculation.^a

Compound	H _f ⁰ calc.	Δ _{calc.−exp.}
Methane	−17.89	0.00
Ethane	−19.74	0.31
Propane	−25.07	−0.05
<i>n</i> -Butane	−30.06	−0.03
<i>n</i> -Pentane	−34.98	0.10
<i>n</i> -Hexane	−39.94	0.01
<i>n</i> -Heptane	−44.90	−0.01
<i>n</i> -Octane	−49.83	−0.01
<i>n</i> -Nonane	−54.76	−0.01
<i>n</i> -Decane	−59.68	−0.05
Isobutane	−32.41	−0.34
Isopentane	−36.72	0.20
Neopentane	−40.86	−0.72
Cyclopentane	−17.90	0.54
Cyclohexane	−29.59	−0.09
Cycloheptane	−28.03	0.19
Cyclooctane	−29.96	−0.23
Cyclononane	−31.41	0.32
Cyclodecane	−37.02	−0.15
Cyclododecane	−53.21	1.38
2,4-Dimethylpentane	−48.11	0.10
3-Ethylpentane	−44.82	0.43
2,2,3-Trimethylbutane	−48.37	0.58
2,2-Dimethylbutane	−44.26	0.22
2,3-Dimethylbutane	−42.02	0.47
3,3-Dimethylpentane	−47.79	0.29
2,2-Dimethylpentane	−49.35	−0.15
Di- <i>t</i> -butylmethane	−57.83	−0.01
3,3-Diethylpentane	−55.73	−0.06
2-Methylhexane	−46.85	−0.33
3-Methylhexane	−45.83	−0.10
2,2-Dimethylhexane	−54.64	−0.96
2,3-Dimethylhexane	−51.20	−0.10
2-Methylheptane	−51.88	−0.47
3-Methylheptane	−50.97	−0.17
4-Methylheptane	−50.95	−0.29
2,2,3,3-Tetramethylbutane	−53.34	0.58

(Continued)

TABLE V.
Continued.

Compound	H _f ⁰ calc.	Δ _{calc.−exp.}
2,2,3,3-Tetramethylpentane	−56.34	0.29
3-Ethyl-3-methylpentane	−50.63	0.72
Methylcyclopentane	−25.68	−0.41
Ethylcyclopentane	−30.25	0.10
1,1-Dimethylcyclopentane	−33.24	−0.20
Norbornane	−12.82	−0.41
<i>cis</i> -Bicyclo[3.3.0]octane	−23.29	−1.22
<i>trans</i> -Bicyclo[3.3.0]octane	−15.69	0.17
Protoadamantane	−20.62	−0.08
eq-Methylcyclohexane	−37.06	−0.07
Bicyclo[2.2.2]octane	−22.07	0.51
Adamantane	−30.84	0.92
<i>trans</i> -Decalin	−43.77	−0.23
<i>cis</i> -Decalin	−40.53	−0.08
Tri- <i>t</i> -butylmethane	−49.43	6.97

Standard deviation = 0.39 kcal / mol; weighted standard deviation = 0.36 kcal / mol.

^aIn kcal / mol. The WT and experimental heats of formation are given in Table III.

we changed the molecular orientation and reran the geometry optimization, and if this did not help, we carried out the calculations again, but using the MM3 starting geometry. In all suspect cases we were able to reduce the error by this procedure. As a test we also applied this procedure to several cases in which no error was suspected, and in those cases the energy change was less than 0.1 kcal/mol.

Results and Discussion

To determine the values for the group and bond equivalents needed to convert the HF energy to a heat of formation, the regression was carried out in the usual way⁷ using the data in Table I as input. The resulting equivalents are listed in Table II for the HF calculations. The calculated heats of formation determined by this method are then given in Table III. Tables IV and V give the corresponding DFT-calculated heat of formation data.

First, let us give a short overview of the results before discussing the details. As expected, the Hartree–Fock calculations gave very reasonable results. We studied 52 diverse hydrocarbons, to represent a good selection of all kinds of saturated structures (excluding only three- and four-membered rings, and molecules containing more than 10 carbons). In each case, we have experimental

data, and also the results of MM4 calculations, with which to compare. The MM4 values agree with experiment for the most part to better than 0.5 kcal/mol. The few larger discrepancies (up to 1.5 kcal/mol) may be a result of experimental problems. Thus, we have reasonable confidence in the experimental/MM4 data.

At the Hartree–Fock level, the rms discrepancies between calculation and experiment for the diverse set of compounds was 0.67 kcal/mol. This rms error is roughly twice the error found with molecular mechanics (MM4) (0.34 kcal/mol over the same set), and the latter number is believed to be experimental error limited. When the DFT calculations were compared with experiment, the rms error was reduced to 0.36 kcal/mol. These numbers immediately lead us to several conclusions. First, the DFT numbers are approximately as good as the molecular mechanics numbers, which generally agree with experiment to within the errors of the experiment. Also, the errors in the Hartree–Fock method are roughly twice as large as those in the DFT or molecular mechanics methods. These larger errors are probably a result of the exclusion of electron correlation in the heats of formation as determined from the HF calculations. That this is indeed the case can be seen in various ways. We start by examining the values of the parameters obtained by the two methods.

It has been known for a very long time that branched-chain hydrocarbons are, generally speaking, more stable than straight-chain analogs. At least some of this additional stability is a result of electron correlation. Branched-chain molecules are more compact, hence the electrons are closer together on average, so that the correlation of their motions lowers the energy more than it would in an extended system. If we look at the parameters that are required in the present calculations, we specifically have methyl, iso, and neo, which are branching arrangements (relative to the ordinary secondary carbon that makes up the bulk of a straight chain and which is the zero of reference here). The methyl increment is slightly positive, iso is negative, and neo is much more negative. These numbers lead to the branch chain stabilization effect in the calculations, and they are expected to be correlation energy dependent. Note that the numerical values for the Hartree–Fock parameters range from 2.6 to -8.5 kcal/mol, a range of 11.1, whereas, in the DFT calculations, where the correlation is more explicitly accounted for (to some approximation), these parameters are

much reduced, from a range of $+1.6$ to -6.8 kcal/mol (a range of 8.4). The difference between the parameter sets is expected to be largely due to the inclusion of some electron correlation in the DFT calculations. The corresponding molecular mechanics parameter set (MM4) spans the range -0.09 to -2.7 or 2.6 kcal/mol, where longer range ($\geq 1,4$) correlation is fully accounted for by van der Waals interactions. And, of course, while correlation can be implicitly included with the aid of these parameters, this is not a highly accurate way to do so, because it would only include correlation between electrons concentrated around atoms arranged 1–2 or 1–3 to each other, but would exclude longer range correlation. Thus, the largest errors in HF calculation should occur when one compares very compact, ball-like molecules with extended chains. The parameterization process used fits the extended chains quite well, for both the HF and DFT calculations. The errors in these calculations span ranges relative to experiment of 1.37 and 0.36 kcal/mol, respectively. Thus, most of the error is pushed into the ball-like molecules when correlation is not included more explicitly. If we look at tri-*t*-butylmethane, which is the most extreme example of such a molecule, we note that the error at the HF level is 13.0 kcal/mol! Also, we note that the DFT calculation reduces this number very much, to 7.0 kcal/mol. The corresponding number from molecular mechanics is 2.3 kcal/mol.

This interpretation of the electron correlation in these molecules is, of course, simplistic, in that experimental errors are present, and some of the correlation is accounted for implicitly in various ways. Nonetheless, we believe that it is correct for the most part. The neglect of correlation in the Hartree–Fock method can be dealt with by parameterization for molecules that are not highly compact; however, as the molecule gets larger, if it is compact, then the neglect of correlation will become extremely serious. The DFT calculation reduces the magnitude of the correlation error by perhaps a factor of 2. However, for molecules larger than about C10, if they are highly compact, they may still contain significant correlation error.

We note that, as groups, the normal open-chain molecules have essentially no van der Waals attraction between remote atoms (opposite ends), but the rings from C6 to C10 do. Thus, the calculated heats of formation of the rings are too high, and those for the open chains too negative at the HF level (Table III). The calculations give much

better agreement with experiment when some correlation is included (DFT and molecular mechanics). The C5 ring also shows this effect. The five-membered ring parameter differs some (-0.67 kcal/mol DFT, -1.38 HF) from the value of the parameter needed for other rings. If it is omitted, the results are conspicuously worse than when it is included. Why this is so is not known with certainty, but it seems likely that this parameter makes up for the electron correlation not included at the level of the present calculations. Such correlation would be expected to lower the DFT energy of a molecule, because the five-membered ring is more compact than larger rings, and it would be expected to lower the energy even more in the HF calculation where correlation is neglected completely. This is exactly what is found, both in sign and magnitude of the corrections. Of course, in a data-fitting procedure, such that used in the present case, any kind of systematic error that appears in five-membered ring structures for any experimental or calculational reason would be absorbed by this parameter. Still, the numbers are such that they appear to be real. Certainly with smaller rings, such parameters are needed in molecular mechanics. We have not included any compounds containing such rings in the present work so it is not known with certainty if they would be required here, but it is expected. The C12 ring is a special case, and it is weighted zero in the parameter fittings. The problem with this compound is that it has a low frequency vibration that cannot be dealt with by the harmonic methods used in the molecular mechanics calculations, and the experimental value is not very certain, in our view.

We may also comment on the value of *TOR*. *TOR* is the excess enthalpy in a torsion with a low barrier, relative to an ordinary vibration, which results because of the lower lying vibrational levels. It is dependent on the barrier height, multiplicity, etc., and is certainly not a constant. However, the work of Pitzer and Gwinn⁸ suggests that a value of about 0.4 kcal/mol for each such vibration in the molecule (excluding rotations of methyl groups already included in the methyl increment) would be a reasonable approximation. In the HF calculation, *TOR* is found to have a value of 1.00 kcal/mol. This suggests that, as an adjustable parameter, it is absorbing some other error or errors as well. But, for the DFT calculations, *TOR* has a value of 0.46, in reasonable agreement with expectations.

Finally, it should be mentioned that the proper way to calculate a heat of formation is to calculate the zero-point and thermal enthalpy of the molecule from the vibrational spectrum, and to include these explicitly in the statistical mechanics calculations.^{9,10} The scheme used here is adapted from the widely used "bond energy" scheme, summarized in Benson's book,² and used in MM2,³ MM3,⁹ and elsewhere. If the full calculation were carried out in the present work, presumably better agreement with experiment should result. The computational effort would, of course, be much greater, and because the disagreement with experiment is already down to approximately experimental error, no real improvement is to be expected. The present method appears to be quite adequate, meaning that the "bond energy" approximations are quite good and, in addition, it keeps the calculations from becoming needlessly complicated and time consuming. The next step is to extend the calculations to functionalized molecules, and obtain important and desired information that is not presently (and probably will never be) available from experimentation.

Conclusions

Heats of formation determined by good experiments, by good molecular mechanics calculations, or by DFT calculations in the present work with the Gaussian program are all of comparable overall accuracy (approximately ± 0.40 kcal/mol over a large set of compounds with one exception for a large and compact molecule). We mentioned earlier that when two of these methods agree, and one disagrees, the one that disagrees is almost certainly in error. We feel even more strongly now that that statement is correct. The Hartree-Fock calculations give numbers that are also comparable with the others, but of less accuracy (± 0.7 kcal/mol). The DFT programs used in the present work normally work well, but perhaps 5% of the time (mostly with larger molecules) they give conspicuously unrealistic results for reasons so far not completely clear. Rerunning the structure optimization from a different starting point in each case has solved the problem. (Rerunning structures for which there is no problem normally gives reproducible results to within < 0.1 kcal/mol for small molecules, or 0.2 kcal/mol for large molecules.)

Acknowledgment

The authors are indebted to Tripos Associates for partial support of this research.

References

1. L. A. Curtis, K. Raghavachari, G. W. Trucks, and J. A. Pople, *J. Am. Chem. Soc.*, **94**, 7221 (1991).
2. S. W. Benson, *Thermochemical Kinetics: Methods for the Estimation of Thermochemical Data and Rate Parameters*, John Wiley & Sons, New York, 1976.
3. N. L. Allinger and U. Burkert, *Molecular Mechanics*, American Chemical Society, Washington, DC, 1982.
4. N. L. Allinger, L. R. Schmitz, I. Motoc, C. Bender, and J. Labanowski, *J. Am. Chem. Soc.*, **114**, 2880 (1992).
5. M. R. Ibrahim and P. v. R. Schleyer, *J. Comput. Chem.*, **6**, 157 (1985).
6. K. B. Wiberg, *J. Comput. Chem.*, **5**, 197 (1984).
7. N. L. Allinger, K. Sakakibara, and J. K. Labanowski, *J. Phys. Chem.*, **99**, 9603 (1995).
8. K. S. Pitzer and W. D. Gwinn, *J. Chem. Phys.*, **10**, 428 (1942). See ref. 3, p. 175, for pertinent discussion.
9. N. L. Allinger, Y. H. Yuh, and J-H. Lii, *J. Am. Chem. Soc.*, **111**, 8551 (1989).
10. N. L. Allinger, K. Chen, and J-H. Lii, *J. Comput. Chem.*, **17**, 642 (1996).
11. (a) M. J. Frisch, G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. W. Wong, J. B. Foresman, M. A. Robb, M. Head-Bordon, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, J. J. P. Stewart, and J. A. Pople, *Gaussian-94/DFT* (Revision E.2), Gaussian Inc., Pittsburgh, PA, 1993. (b) A. D. Becke, *J. Phys. Chem.*, **98**, 5648 (1993). (c) C. Lee, W. Yang, and R. G. Parr, *Phys. Rev. B*, **37**, 785 (1988).
12. (a) S. H. Vosko, L. Wilk, and M. Nusair, *Can. J. Phys.*, **58**, 1200 (1980); (b) J. P. Perdew, *Phys. Rev.*, **B33**, 8822; see also erratum: **38**, 7406 (1986); (c) A. D. Becke, *Phys. Rev.*, **A38**, 3098 (1988); (d) D. R. Salahub, R. Fournier, P. Mlynarski, I. Papai, A. St-Amant, and J. Ushio, In *Density Functional Methods in Chemistry*, J. Labanowski and J. Andzelm, Eds. Springer, New York, 1991, p. 77.